

Survival Chances of Mutants Starting With One Individual

CHRISTOPH KUHN

*Institut für Molekularbiologie und Biophysik, Gruppe Biophysik, ETH-Hönggerberg, CH-8093
Zürich, Switzerland, e-mail: c-k@gmx.ch*

Abstract. A simple theoretical model of a Darwinian system (a periodic system with a multiplication phase and a selection phase) of entities (initial form of polymer strand, primary mutant and satellite mutants) is given.

First case: one mutant is considered. One individual of the mutant appears in the multiplication phase of the first generation. The probabilities to find N mutants $W_n^M(N)$ after the multiplication phase M of the n -th generation (with probability δ of an error in the replication, where all possible errors are fatal errors) and $W_n^S(N)$ after the following selection phase S (with probability β that one individual survives) are given iteratively. The evolutionary tree is evaluated. Averages from the distributions and the probability of extinction $W_\infty^S(0)$ are obtained.

Second case: two mutants are considered (primary mutant and new form). One individual of the primary mutant appears in the multiplication phase of the first generation. The probabilities to find N_p primary mutants and N_m of the new form $W_n^M(N_p, N_m)$ after the multiplication phase M of the n -th generation (probability ε of an error in the replication of the primary mutant giving the new form) and $W_n^S(N_p, N_m)$ after the following selection phase S (probabilities β_p and β_m that one individual each of the primary mutant and of the new form survives) are given iteratively. Again the evolutionary tree is evaluated. Averages from the distributions are obtained.

Key words: Darwinian system, multiplication, error in replication, selection, initial form, primary mutant, satellite mutants, Bernoulli polynomial, probability distribution, evolutionary tree, complementary and anti-parallel replication, plus-strand, minus-strand, translation apparatus, origin of life

1. Introduction: Darwinian System

1.1. AVERAGE NUMBER OF MUTANTS

In a simple theoretical model of a Darwinian system proposed by the present author and others [1–4], there is a multiplication phase M and selection phase S (Figure 1) of a population of entities (e.g. polymer strands of different folding structures due to different sequences). In the multiplication phase the strands replicate by multiplication factor ρ (with the number S of subsequent replication steps, $\rho = 2^S$). In the selection phase, a single strand survives with probability β . The multiplication factor per generation is $r = \rho\beta$.

For the initial form let us suppose $r = 1$, then the total population stays constant. One individual of the primary mutant appears with $r > 1$, where the surviving

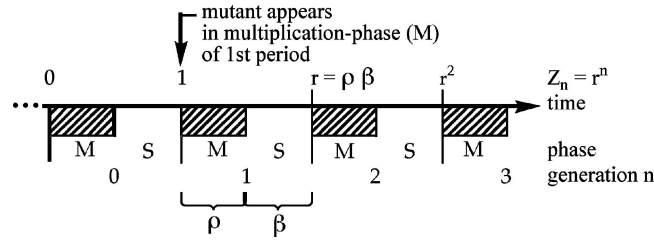


Figure 1. Darwinian system with a multiplication phase M and a selection phase S of an entity (polymer strand). One mutant appears at the beginning of first generation. The average number of mutants after n generations, if supply is not the limiting factor, is Z_n . The multiplication factor is ρ , and β is the probability that an individual of the mutant survives during selection phase.

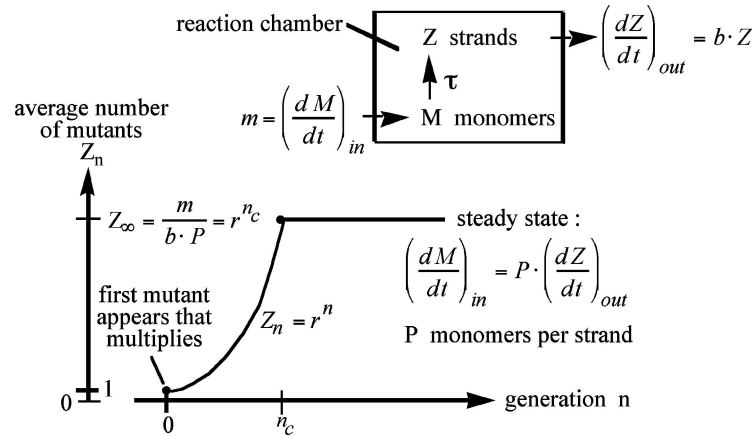


Figure 2. Cut-off Z_∞ of the average number of mutants, when the limit of supply is reached (steady state in reaction chamber).

probability β for the primary mutant is larger than for the initial form. Thus the average number of primary mutants is $Z_n = r^n$ after n generations if supply is not the limiting factor. There must be a cut-off when the limit of supply is reached (Figure 2).

1.2. PRIMARY MUTANT AND SATELLITE MUTANTS

Consider the average number Z of primary (error-free) mutants and Z_{kF} of satellite (k -error containing) mutants (Figure 3). This state is reached from one generation to the next by replication either with one additional error in a copy of a $(k - 1)$ -error containing mutant (factor $(\rho - 1) \eta$ in first approximation, diagonal paths in Figure 3) or with no additional error in a copy of a k -error containing mutant (factor $\rho - (\rho - 1) \eta$ in first approximation, horizontal paths in Figure 3). The

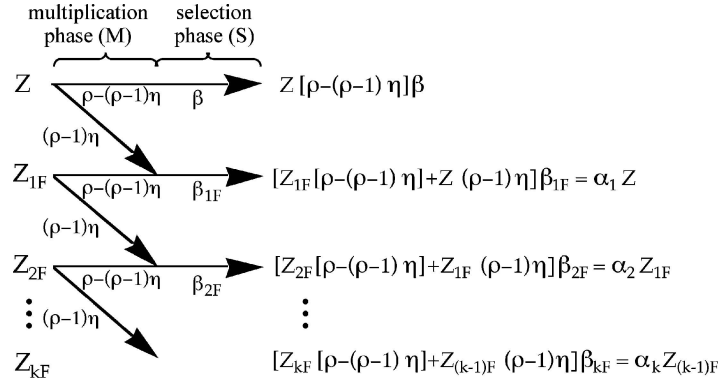


Figure 3. Average number Z of a primary (error-free) mutant and Z_{kF} of satellite (k -errors containing) mutants. Diagonal paths: one additional error in a copy of a $(k - 1)$ -error containing mutant. Horizontal paths: no additional error in a copy of k -error containing mutant. Probability η that an error occurs in a given strand. Selection of mutants occurs with probability $\beta > \beta_{1F} > \beta_{2F} > \dots > \beta_{kF}$. Steady state for values $Z, Z_{1F}, Z_{2F}, \dots, Z_{kF}$.

probability that an error occurs in a given strand is given by η . Selection of mutants occurs with probability $\beta > \beta_{1F} > \beta_{2F} > \dots > \beta_{kF}$. Considering the steady state, the values $Z, Z_{1F}, Z_{2F}, \dots, Z_{kF}$ from one generation (left side in Figure 3) to the next (right side in Figure 3) are identified; note that $[\rho - (\rho - 1) \eta] \beta = 1$ and $\alpha_k = (\rho - 1) \eta \beta_{kF} / [1 - (\rho - (\rho - 1) \eta) \beta_{kF}]$. Assuming $\alpha_k = \alpha / k$ for simplicity, then $Z_{kF} = (\alpha / k) Z_{(k-1)F} = (\alpha_k / k!) Z$, and with $Z_{\text{total}} = Z + Z_{1F} + Z_{2F} + \dots = Z e^\alpha$ we find

$$Z_{kF} / Z_{\text{total}} = (\alpha_k / k!) / e^\alpha \quad (1)$$

and this result is shown graphically in Figure 4.

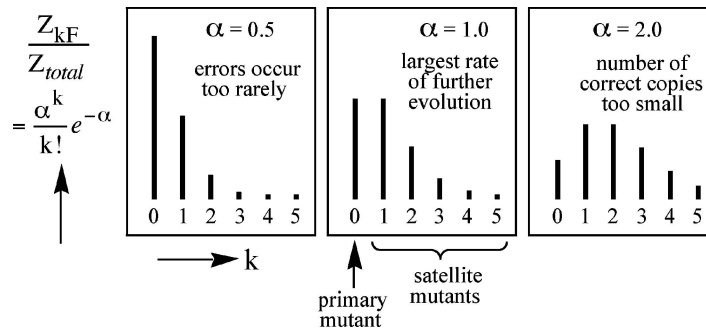


Figure 4. Primary (error-free) mutant and satellite (k -error containing) mutants. The corresponding average numbers are Z and Z_{kF} .

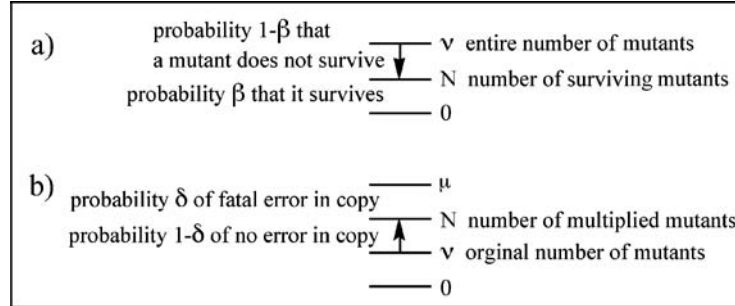


Figure 5. Sketch of how the Bernoulli polynomial is applied to the probability distribution for the case of a single mutant. (a) Selection phase, Equation (2). (b) Multiplication phase, Equation (3). $(\mu - v)$ total copies of mutant, $(N - v)$ error-free copies of mutant. All possible errors are fatal errors.

2. Probability Distribution of N Mutants

2.1. ONE MUTANT CONSIDERED

The probability to find $N \leq N_{\max}$ number of mutants after the selection phase S (dots (•) in the evolutionary trees in Figures 6 and 7) of the n -th generation is given by the Bernoulli polynomial (Figure 5a)

$$W_n^S(N) = \sum_{v=N}^N W_n^M(v) \binom{v}{N} (1 - \beta)^{(v-N)} \beta^N \quad (2)$$

where β is the probability that one mutant individual survives.

The probability to find $N (\leq N_{\max})$ number of mutants after the multiplication phase M (circles (○) in the evolutionary trees in Figures 6 and 7) of the $(n + 1)$ -th generation is given by the Bernoulli polynomial (Figure 5b)

$$W_{n+1}^M(N) = \sum_{v=L(N)}^N W_n^S(v) \binom{\mu - v}{N - v} \delta^{(\mu-N)} (1 - \delta)^{(N-v)} \quad (3)$$

where δ is the probability that a copy of the mutant has a fatal error, where all possible errors are fatal errors, $\mu = \begin{cases} 2v & 2v \leq N_{\max} \\ N_{\max} & 2v > N_{\max} \end{cases}$ the cut-off condition when the limit of supply is reached and $L(N) = \begin{cases} N/2 & \text{even } N \\ (N+1)/2 & \text{odd } N \end{cases}$.

The evolution trees (Figures 6 and 7) are given by the initial probability $W_{\text{initial}}(1) = 1$ that one individual of the mutant appears: $W_1^M(1) = W_{\text{initial}}(1)\delta = \delta$ and $W_1^M(2) = W_{\text{initial}}(1)(1 - \delta) = (1 - \delta)$.

The numerical evaluations are given in Figures 8–10. The average number Z_n of mutants is evaluated from the probability distribution (Figures 8 and 9) for the r^n -regime ($n = 9$) and for the Z_∞ -regime ($n = 30$). The probability of extinction $W_\infty^S(0)$ as function of δ and β is given in Figure 10. With three subsequent

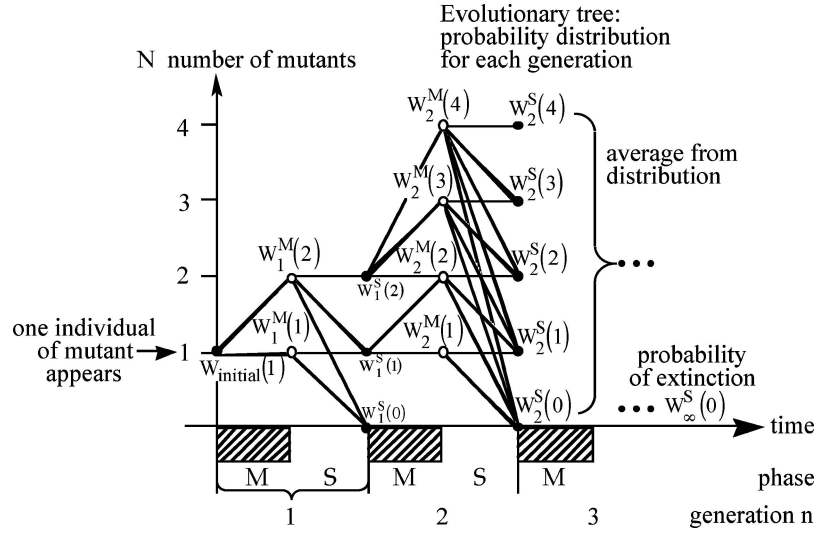


Figure 6. Evolutionary tree of mutants. Probability distributions (Equations (3) and (2)) of finding N mutants after the multiplication phase (\circ) and after the selection phase (\bullet) for each generation starting with one mutant individual. The probability of extinction is $W_\infty^S(0)$.

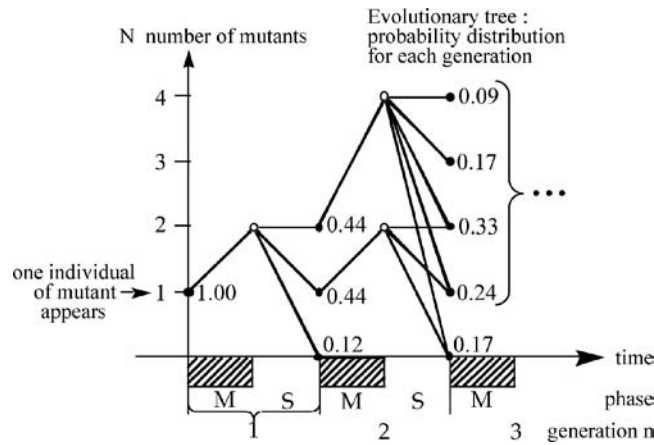


Figure 7. Evolutionary tree for the simple case where: $\delta = 0$ (no errors occurring), $\beta = 2/3$ and $\rho = 2$ (one replication step in multiplication phase).

replication steps in each multiplication phase (Figure 10b) the multiplication factor $\rho = 8$ and smaller β values become accessible.

2.2. TWO MUTANTS CONSIDERED

The probability to find N_p primary mutants and N_m of the new form ($N_p + N_m = N \leq N_{\max}$) after the selection phase S of the n -th generation is given by the

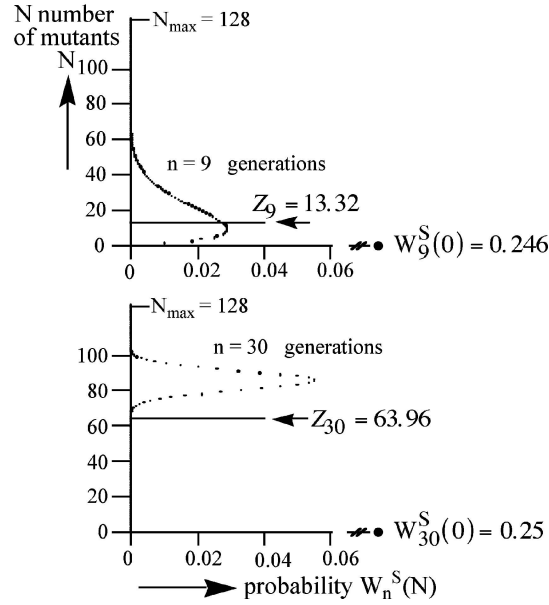


Figure 8. Probability distributions (continued from Figure 7) of finding N mutants after the selection phase at generations $n = 9$ and $n = 30$. We assume a limited monomer supply ($N_{\max} = 2^7 = 128$).

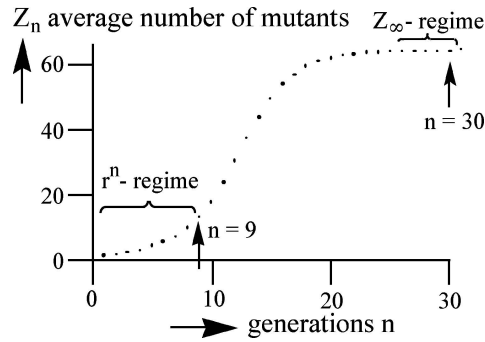


Figure 9. Average number Z_n of mutants evaluated from probability distribution (note that the probability of extinction $W_n^S(0)$ must be taken into account). The two cardinal examples (Figure 8) $n = 9$ (r^n -regime: $r = \rho\beta = 4/3$, $(4/3)^9 = 13.32$) and $n = 30$ (Z_∞ -regime: $Z_\infty = 64.0$) are indicated by arrows.

Bernoulli polynomial (Figure 11a)

$$W_n^S(N_p, N_m) = \sum_{v_p=N_p}^q \sum_{v_m=N_m}^q W_n^M(v_p, v_m) \cdot \binom{v_p}{N_p} (1 - \beta_p)^{(v_p - N_p)} \beta_p^{N_p} \cdot \binom{v_m}{N_m} (1 - \beta_m)^{(v_m - N_m)} \beta_m^{N_m} \quad (4)$$

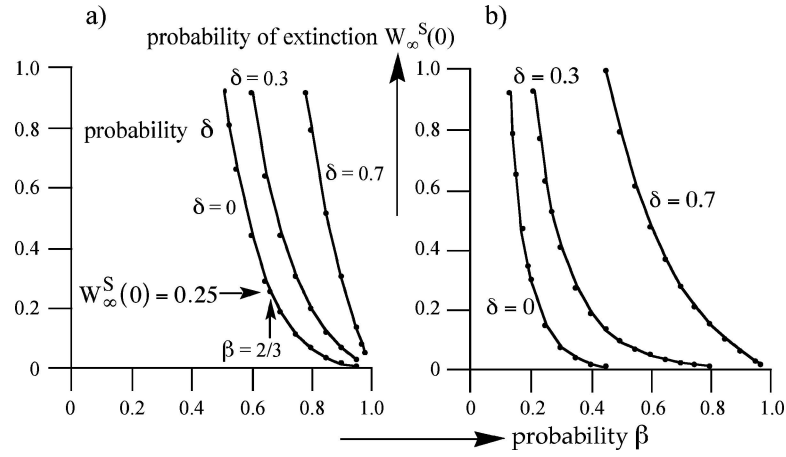


Figure 10. Probability of extinction $W_\infty^S(0)$ as a function of δ and β . (a) One subsequent replication step in each multiplication phase ($\rho = 2$). The case of Figures 7–9 ($\delta = 0$ and $\beta = 2/3$, $W_\infty^S(0) = 0.25$) is indicated by arrows. (b) Three subsequent replication steps in each multiplication phase ($\rho = 8$).

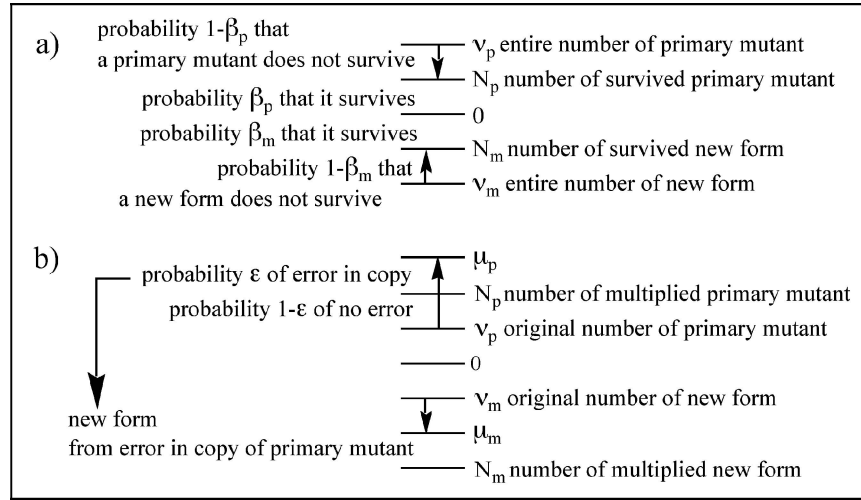


Figure 11. Sketch of how the Bernoulli polynomial is applied to the probability distribution. (a) Selection phase, Equation (4). (b) Multiplication phase, Equation (5). For the case of two mutants, there are $(\mu_p - v_p)$ total copies of the primary mutant, $(N_p - v_p)$ error-free copies of the mutant, $(\mu_m - v_m)$ total copies of the new form, and $(\mu_p - N_p)$ error-containing copies of the primary mutant which give the new form.

where β_p (β_m) is the probability that one individual of the primary mutant (of the new form) survives.

The probability to find N_p primary mutants and N_m number of the new form ($N_p + N_m = N \leq N_{\max}$) after the multiplication phase M of the $(n+1)$ -th generation

is given by the Bernoulli polynomial (Figure 11b)

$$W_{n+1}^M(N_p, N_m) = \sum_{v_p=0}^v \sum_{v_m=0}^{v-v_p} \sum_{g_p=0}^{\mu_p-v_p} W_n^S(v_p, v_m) \cdot \binom{\mu_p-v_p}{g_p} \varepsilon^{(g_p)} (1-\varepsilon)^{(\mu_p-g_p-v_p)} \quad (5)$$

where ε is the probability that an error in copying the primary mutant occurs and the new form emerges and, $\mu_p = \begin{cases} 2v_p & 2v_p \leq N_{\max} \\ (v_p/v)N_{\max} & 2v_p > N_{\max} \end{cases}$ and $\mu_m = \begin{cases} 2v_m & 2v_m \leq N_{\max} \\ (v_m/v)N_{\max} & 2v_m > N_{\max} \end{cases}$ are the cut-off conditions when the limit of supply is reached.

The evolutionary tree is given by the initial probability $W_{\text{initial}}(N_p = 1, N_m = 0) = 1$ that one individual of the primary mutant appears: $W_1^M(N_p = 2, N_m = 0) = (1 - \varepsilon)W_{\text{initial}}(1, 0) = 1 - \varepsilon$ and $W_1^M(N_p = 1, N_m = 1) = \varepsilon W_{\text{initial}}(1, 0) = \varepsilon$. The

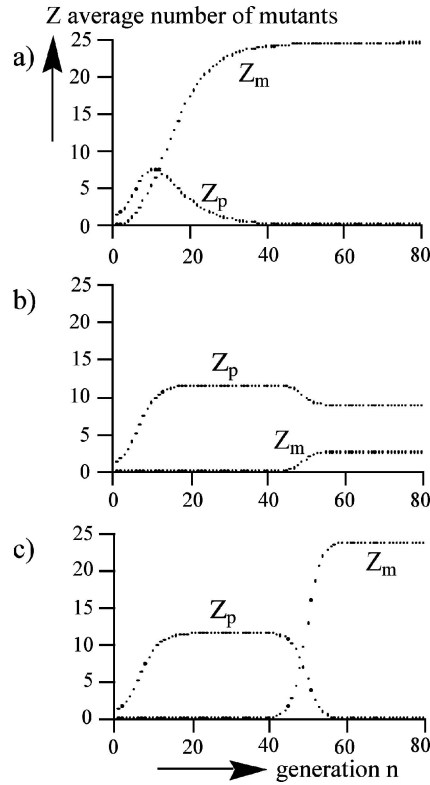


Figure 12. Average number of mutants Z_p and Z_m evaluated from the probability distribution. There are two subsequent replication steps in each multiplication phase ($\rho = 4$), and we assume limited monomer supply ($N_{\max} = 2^6 = 64$). (a) Probability of error resulting in new form ($\varepsilon = 0.005$), probability that primary mutant and new form survives ($\beta_p = 0.33$ and $\beta_m = 0.67$). (b) and (c) $\varepsilon = 0.05 \operatorname{sech} [(n - 50)/2]$ (smooth peak at $n = 50 \pm 5$), $\beta_p = 0.33$. (b) $\beta_m = 0.33$. (c) $\beta_m = 0.67$.

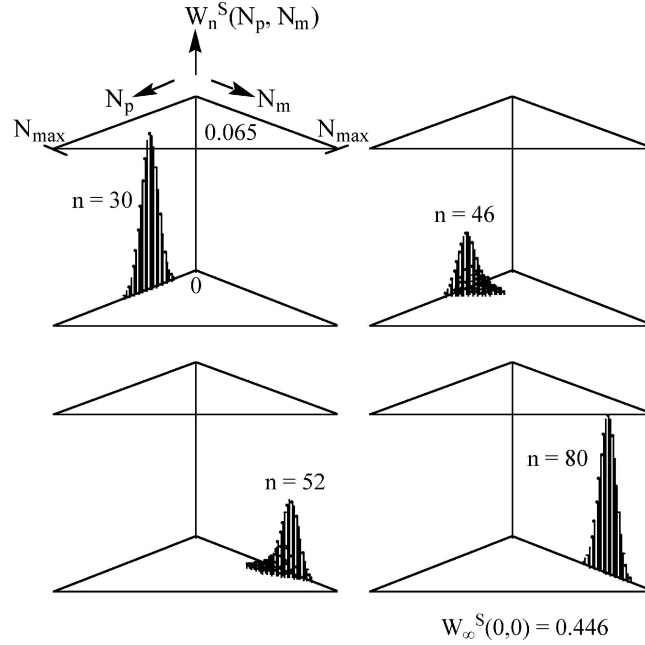


Figure 13. Probability distribution $W_n^S(N_p, N_m)$ of finding N_p primary mutants and N_m of the new form ($N_p + N_m = N \leq N_{\max} = 64$) after the selection phase (cases from Figure 12c) in the n -th generation (indicated).

formalism can be applied to systems with time-dependent ε (changing exposition to mutagen): in Figure 12a ε is assumed to be time-independent, in Figures 12b, 12c and 13 ε is chosen to have a smooth peak at $n = 50 \pm 5$.

2.3. ELIMINATION OF THE (−)-R-STRAND, RETENTION OF THE (+)-R-STRAND: A CRUCIAL STEP IN ORIGIN OF LIFE

Starting from a “RNA-world” (R -strands building Hairpin-Assembler devices) and introducing A -monomers (which interlink to the HA-devices such that they become bound and form an A -oligomer) it is shown that a simple translation apparatus evolves (HAE-device producing enzymes [1–4]). In this “RNA-protein world” the complementary and anti-parallel copy of the (+)-assembler (carrying the recipe to produce an enzyme) is a (−)-assembler (its HAE-device producing nonsensical agglutinate).

Later on D -strands are introduced, which are not capable of participating in the synthesis of A -oligomers. In the “DNA-RNA-protein world” the mixture of (+)-assembler and (−)-assembler is unstable eliminating (−)-assembler: (+)- R -strands having (−)- D -strands as template and (−)- D -strands having (+)- R -strands

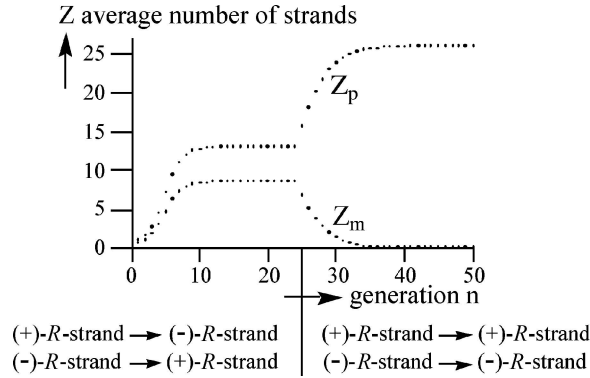


Figure 14. Average number of mutants Z_p and Z_m evaluated from the probability distribution. Probability that (+)-R-strand and (-)-R-strand survives are $\beta_p = 0.33$ and $\beta_m = 0.50$ respectively; we consider two subsequent replication steps in each multiplication phase ($\rho = 4$), and assume limited monomer supply ($N_{\max} = 64$).

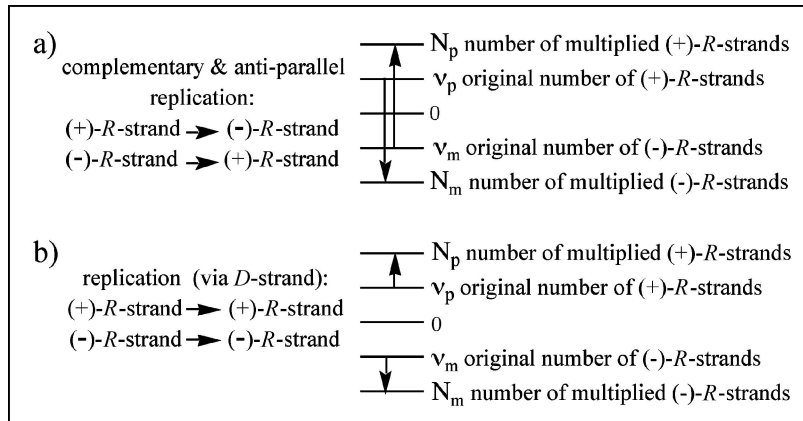


Figure 15. (a) Average number of mutants Z_p and Z_m evaluated from probability distribution, (b) probability distribution $W_n^S(N_p, N_m)$ of finding N_p number of (+)-R-strands and N_m number of (-)-R-strands after the selection phase in the n -th generation (indicated) before and after the transition. Transition at generation $n = 25$. Probability that (+)-R-strand and (-)-R-strand survives ($\beta_p = 0.33$ and $\beta_m = 0.50$), two subsequent replication steps in each multiplication phase ($\rho = 4$), limited monomer supply ($N_{\max} = 64$).

as template will be selected. These (-)-D-strands become carriers of the genetic information.

This elimination of the (-)-assembler is treated simply as in the following (the numerical evaluation is given in Figure 14). For the selection phase we use Equation (4) with the probability $\beta_p > \beta_m$ that a (+)-R-strand or a (-)-R-strand survives, respectively (Figure 11a). For the multiplication phase in the “RNA-protein

world” we use the replication rule (Figure 15a):

$$\begin{aligned} N_p &= v_p + v_m \\ N_m &= v_m + v_p. \end{aligned} \tag{6}$$

In the “DNA-RNA-protein world” we use the replication rule (Figure 15b):

$$\begin{aligned} N_p &= 2v_p \\ N_m &= 2v_m. \end{aligned} \tag{7}$$

3. Conclusion

A simple theoretical model of a Darwinian system is given. Probability distributions for replication and for selection (Bernoulli polynomials) are applied iteratively. The cases of only one mutant and of two mutants (primary mutant and new form) are considered. The formalism is applied to a crucial step in the origin of life: the elimination of the (–)-*R*-strand and retention of the (+)-*R*-strand.

References

1. Kuhn, H. and Waser, J.: Self-Organization of Matter and Early Evolution of Life, in W. Hoppe, W. Lohmann, H. Markl and H. Ziegler (eds.), *Biophysics*, Springer-Verlag, Berlin, 1983, pp. 830–874.
2. Kuhn, H. and Waser, J.: A Model of the Origin of Life and Perspectives in Supramolecular Engineering, in J. -P. Behr (ed.), *Lock-and-key principle*, Wiley, Chichester, 1994, pp. 247–306.
3. Kuhn, C.: Computer-Modeling Origin of a Simple Genetic Apparatus, *Proceedings of the National Academy of Sciences USA* **98** (2001), 8620–8626.
4. Kuhn, H. and Kuhn, C. Diversified World: Drive of Life’s Origin?! *Angewadte Chemical International Edition* **42** (2003), 262–266.